

Heart Rate Regulation processed through wavelet analysis and change detection. Some case studies

Nadia KHALFA [†] Pierre, R. BERTRAND^{*†}
Gil BOUDET [§] Alain CHAMOUX [¶] and Véronique BILLAT [#]
[†] *ENIT, Tunisia and INRIA Saclay, France*
E-mail: nadia.khalfa@inria.fr
^{††} *INRIA Saclay and Clermont University, France.*
E-mail: Pierre.Bertrand@inria.fr
[§] *CHU Clermont Ferrand*
E-mail: gil.boudet@u-clermont1.fr
[¶] *CHU Clermont Ferrand*
E-mail: alain.chamoux@u-clermont1.fr
[#] *UBIAE, Inserm Team U.902 and Évry Génopole*
E-mail: veronique.billat@wanadoo.fr

December 29, 2010

Abstract

Heart rate variability (HRV) is an indicator of the regulation of the heart engine, Task Force (1996). This study compares the regulation of the heart in two cases of healthy subjects within ecological situations: Marathon runners and shift workers. After an update on the state of the art on HRV processing, we specify our probabilistic model: We choose modeling heartbeat series by locally stationary Gaussian process (Dalhaus, 1997). HRV and heart regulation are then processed by the combination of two statistical methods: Continuous wavelet transform for calculating the spectral density energy in the high frequency (HF) and low frequency (LF) bands and change point analysis to detect changes of heart regulation. Next, we plot the variations of the HF and LF energy in extreme conditions for both populations. This put in light, that physical activities (rest, moderate sport, marathon race) can be ordered in a logical continuum. The results obtained are relevant but have to be completed by further studies.

Keywords: Heart Rate Variability; Ecological conditions; Locally stationary Gaussian process; Spectral analysis; Continuous wavelet transform; Change points analysis; Index of heart regulation.

^{*}corresponding author: Pierre, R. Bertrand, INRIA Centre de recherche Saclay - Île-de France, Parc Orsay Université, 4, rue Jacques Monod 91893 ORSAY Cedex France.

Introduction

Physical activities induce variation of the heart rate (HR), which is defined as the number of heartbeats by minute. However, devices are not measuring heart rate but the interval between two successive R-waves. The RR interval corresponds to the duration (in seconds or milliseconds) of each single heartbeat. The two quantities are linked by the equation $HR = 60/RR$, and in this work, we will use indifferently both notations. Except in some very severe diseases, heartbeat series is fluctuating. This fact is known as heart rate variability (HRV). HRV has retained the attention of cardiologists, see Task Force (1996) or Huikuri *et al.*(2009) for a recent survey. Indeed, HRV reflects the regulation of HR and can contain informations not visible but relevant.

A hidden drawback is that a heartbeat series is a stochastic time dependent signal and almost also fluctuating and complicated than a financial process. One novelty of this work is the use of theoretical probabilistic model combined with statistics for heartbeats series.

During the two last decades, many studies of HRV have been done. Some forget the dynamic structure of heartbeat series. Others analyze the dynamic structure of heartbeat series : A first family of method was introduced during the 1990's and is based on spectral analysis. It consists on the computation of the spectral density in different bands of frequency by using the Fourier transform or the wavelet transform for example. An apparently alternative set of methods is based on fractal analysis, see Ivanov *et al.* (1999) or Huikuri *et al.*(2009) and the references therein.

All methods assume, at least implicitly, that a heartbeat series is a stochastic process X depending on a d -dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, where Θ is a subset of the whole space. The parameter θ , which varies depending on the experimental conditions, is assumed to provide relevant informations on HR or HR regulation. Saying that the process X , also denote by X_θ , depends on parameter θ is equivalent to say that X_θ belongs to a certain class of models $\mathcal{M} = \{X_\theta, \theta \in \Theta\}$. Next, the logical question becomes: Which stochastic model shall we use for heartbeats series? As pointed out by G. Box: "*All models are wrong, but some are useful*", and particularly simple models are easier and therefore more useful. In care centers, hospitals or laboratories, experimental conditions are under control and can be assumed to stay fixed. As a corollary, the environment remains independent of time and a mathematical translation is furnished by stationary process, that is process which structural parameters and time dependency structure stay fixed during time. In order to remain simple, we will also assume that the process is Gaussian.

A second novelty of this work is the use of small devices outside hospital, care centers or laboratories. We dealt with heartbeat series measured in so-called ecological situations: Datasets of shift workers furnished by Clermont-Ferrand hospital and datasets of Paris marathon runners furnished by Inserm Team U. 902. In these cases, we implicitly assume that the structural parameters of the heartbeats series would vary depending on the environmental or physiological condition, that is to say following the kind of activity of the shift worker or onset of disease at the end of a marathon race. The most simple stochastic model is furnished by locally stationary Gaussian process, introduced by Dalhaus (1997). "*Locally*" means the existence of a segmentation $\tau = (\tau_1, \dots, \tau_K)$ such that the process X is a Gaussian stationary process on each sub-interval (τ_k, τ_{k+1}) for $k = 0, \dots, K$, where by convention τ_0 and τ_{K+1} are respectively the initial time and the final time. Stress that K is the number of changes and can be equal to zero.

From a statistical point of view, this modelling addresses two sets of questions: estimating the

structural parameters θ of stationary Gaussian process and change point analysis of these parameters. Recent technological progress allows recording of heartbeats series of large size (40,000 or 120,000) which addresses a third challenge: Process heartbeats series fast enough. Reasonable solutions have been proposed in Bertrand *et al.* (2011) and Ayache & Bertrand (2011), and we use them in this work. To put it into a nutshell, we compute the spectral density using continuous wavelet transform and then we apply the fast segmentation method so-called FDpV introduced in Bertrand *et al.* (2011).

Going outside hospital and processing heartbeat series of healthy subject in ecological situations leads to slightly different questions. Rather than severe diseases and vital prognostic, we are then concerned by well-being, stress or physiological tiredness, see *e.g.* Billat *et al.* (2009). Our long term objective is building a physiological index of well-being. A first step in this direction is proposed in this work, based on a linear combination of HF and LF wavelet energy segmented by FDpV method. With this new index, we can see that activities of moderate intensity (professional context) and vigorous activities (sport context) are ordered in a logical continuum.

In the rest of this paper, our plan will be the following: In the first section, we describe the experimental protocol. In Section 2, we make an overview of the state of the art for modelling and processing heartbeat series. In the same section, we introduce our probabilistic models for that type of time series, both in the stationary case and in the locally stationary case. In Section 3, we summarize the results obtained by the application of our technologies on the data, completed by a short statistical study. Eventually, in Section 4 we made a discussion on the biological meaning of these first results, and we propose future study perspectives in the conclusion.

1 Data and experiment

The experiment of the present study was held within two different populations carrying different activities. In this section, we introduce these populations and then describe the data acquisition and the pre-processing techniques.

1.1 Subjects

This study concerns long distance Paris marathon runners¹ and shift workers of a company Z that we can not name for lack of permission². The marathon of Paris is 42,195 km long. It lasts from 3 hours to 4 hours and 30 minutes and plus. Runners need to get a medical certificate affirming their physical fitness before participating in the race. The marathon race is no longer reserved to elite runners. For the greatest marathon (Boston, New York, Paris, London), the average performance is 3h50min and the average marathon age is 45 years (both for the females and males) with a standard deviation of 4 years. The marathon race is run at a high percentage (80 – 100%) of maximal heart rate reserve and oxygen uptake, see for example Chamoux (1984) or Cramér and Leadbetter (1967).

Shift workers are in charge of several activities. It is mainly divided into two parts: Loading and unloading of bulk. Loading activity occurs in the evening from 5 pm to 9 pm while the unloading of bulk

¹These recordings were supplied by the laboratory accredited by Inserm and headed by Véronique Billat: Unit 902-Genopole-Université d'Evry-Val-d'Essonne - "Integrative Biology of adaptation to exercise"

²Data furnished by Gil Boudet and Alain Chamoux from the Institute of Industrial Hygiene, occupational medicine and ergonomics, Faculty of Medicine of Clermont-Ferrand

is done in the morning from 5 am to 12 pm. The activity is very repetitive in both cases: lifting, carrying, transporting and depositing loads. Hence, the concern is about the quantification of the harshness of company Z's shift workers' activities. We are convinced that an analysis of the heartbeat time series will respond to the latter question.

1.2 Data acquisition and pre-processing

Marathon runners

For marathon race, the studied cohort is composed of ten non elite marathon runners. Before participation, all subjects were informed of the risks and stresses associated with the protocol, and gave their written voluntary informed consent. The present study conformed to the standards set by the Declaration of Helsinki and its procedures were approved by the local ethics committee of the Saint Louis Hospital of Paris. Heart rate signal was recorded during Paris marathon in April 2006 using a cardio-frequency meter (Polar RS 800, FI). The race commenced at 8:45 am and a temperature of 17 C without wind (< 2 m/s, anemometer, Windwatch, ALBA, Silva, Sweden) was observed. Two weeks prior to the race, the subjects performed a test to determine individual VO_2max and $HRmax$.

Shift workers

For shift workers, measures are made using Spiderview of ELA Medical (now Sorine Group), a very compact numeric Holter recorder. It runs on batteries and can make recordings of 24 hours continuously at the workplace. It is worn on a belt or shoulder, and is not cumbersome. Its 5 adhesive electrodes receive the signal by two derivations and insure reliable recording. The data are stored on memory card. High-resolution ECGs (1000Hz, 2.5 uV) are obtained.

Preprocessing

Heart rate recording during exercise is not an easy work because it is realized in "free-running" conditions and not within controlled clinical conditions. Indeed, bad contact between the worker's skin and the frequency meter added to the possible bad manipulation of the device may induce the presence of aberrant data.

It is worth noting that the size of our data series ranges from 30,000 (for marathon runners) to 150,000 (for shift workers), so it would take much time to carry on all the changes and corrections in order to get an exploitable series of data for further study. While focusing on the previous works, we found that most of them make a part of the corrections of what they call "ectopic beats" Cysarz *et al.* (2008), "premature beats" or noise Pikkujämsä *et al.* (2001) manually with the help of an "experienced observer" Singh *et al.* (1996). So, the use of these techniques is different from a person to an other and is not practical for huge data sets. That being said, it exists programs used to analyze such recordings in hospitals which are implemented for sick people; consequently interpretations can be inappropriate or wrong in case of healthy subjects. For example, for a high level athlete, the interval between two heartbeats can easily reach 450 ms (133 bpm) which is abnormal in the case of people lying in hospital.

For that purpose, and by referring to physiological arguments, we developed a technique named

“Tachogram cleaning”³ capable to automatically identify artifacts and correct them in healthy heartbeat series.

As a by-product, we can automatically measure the rate of aberrant data and then correct them. The average rate of aberrant data was around 7% for heartbeats series of marathon runner and around 1% for shift workers. Following the standard, see Task Force (1996) or Cysarz *et al.* (2008), we reject datasets with a rate of aberrant data greater than 10%. By doing so, 7 datasets of marathon runners were retained in this study and 11 datasets of shift workers. However, only 5 shift workers have intense physical activity outside working hours, that is corresponding to heart rate greater than 130 beats/minute during at least 10 minutes.

2 Method

In this section, we first give an overview on two approaches used in literature on the study of HRV inspired from Task Force (1996) and Huikuri *et al.* (2008). Most of the methods suppose the stationarity of the signal like time and frequency domain ones and others in addition to stationarity consider the non-linear character of the heart rate namely for the fractal approach.

The two first subsections summarize the statistical state of the art of the two used static methods to quantify HRV. In the third subsection, we go back to more mathematical modeling and by doing so we can explain that both methods are linked. Eventually, in the fourth subsection, we propose a new method of analysis for quantifying HRV.

2.1 The state of the art

Static approach

Heart rate variability has been evaluated by an important number of methods supposing the static character of heartbeat time series. The simplest and most popular techniques for studying HRV are time domain measurements that can be divided into statistical and geometric methods.

Statistical approach is particularly applied for long period heartbeat time series (24 h) and calculated rather from direct measurements of RR intervals or differences between them. As a case in point, SDNN is the standard deviation of a series of RR intervals; when it is calculated throughout 24-hour period, it encompasses High frequency (HF) and Low Frequency (LF) components of the series. SDNN increases with the length of the recording; thus it can not be a robust statistical indicator. For other measures derived from RR interval differences, we can note RMSSD, the square root of the mean squared differences of successive RR intervals or NN50, the number of interval differences greater than 50 ms. On the other hand, the geometric approach is based on the conversion of the series of RR intervals into a geometric pattern such as Poincaré Plot (D’Addio *et al.* (2004)) and Lorenz plot which is used in a very recent study of Tsuboi *et al.* (2010)

This being said, the most widely used methods to assess HRV are frequency domain methods. Indeed, heartbeat signal is a time series, thus we can note another structural parameter which is time dependence or correlation structure. This correlation structure varies following the different scales of time. This could be analyzed easily through Fourier Transform. Following recommendations of the *Task Force* (1996),

³ “Tachogram cleaning” has been developed by Nadia Khalfa and Pierre R. Bertrand at INRIA in 2009. This software is deposited.

we consider Fourier Transform of tachogram and the corresponding energy in Very Low Frequency, Low Frequency and High Frequency bands, denoted in the sequel by VLF, LF, HF. The VLF corresponds to frequencies under 0.03Hz, the LF corresponds to the frequency band [0.04Hz, 0.15Hz] and the HF deals with frequencies of the frequency band [0.15Hz, 0.4Hz]. Actually, the state of the art in cardiology is to consider the energy of Fourier Transform into the three frequency bands VLF, LF, HF, and the ratio HF/LF. These different frequency bands are interpreted as referring to different regulation systems: the VLF have been attributed to the renin-angiotensin system, other humoral factors and thermoregulation, the LF is supposed to reflect the activity of the two components of the autonomous nervous system the orthosympathetic and parasympathetic while the HF band is the response of the parasympathetic activity.

Many studies have shown the inability of these methods known as traditional measures of HRV to detect subtle changes in RR series (see Picho *et al.*(1999)). Therefore, recent methods based on the chaos theory and the fractal behaviour have been used to assess HR dynamics.

Non-linear methods: Fractal and complexity measures

Several studies suggested that “HR dynamics are induced by non-linear phenomena”, see *e.g.* Heikki *et al.* (2009) or Bardet and Kammoun (2008), and that is why non-linear measures of HR dynamics are even better predictors of mortality in certain diseases compared to the traditional measures.

For fractal measures, we can cite the power-law HRV analysis and the detrended fluctuation analysis (DFA). The first method is about a plot of spectral power and frequency on a bi-logarithmic scale showing linear portion between 10^{-4} and 10^{-2} Hz with a slope of -1 for healthy subjects and lower values with advancing age. This slope reflects long term scaling characteristics of heart rate fluctuations in the region of the ultralow and very-low frequency bands. The detrended fluctuation analysis (DFA), introduced in Peng *et al.* (1994), estimates the scaling exponents of the detrended integrated time series. In Ivanov *et al.* (1999) and Ching and Tsang (2007), fractal analysis has suggested the fractal character of the heartbeat variations in healthy subjects and the loss of it in case of pathological states.

Billat *et al.* (2009) showed a change in short term scaling exponent of HR calculated by DFA in the second half of the marathon race. This finding provide evidence of the significant effect of fatigue induced by long exercise on HR and the efficiency of DFA. Meanwhile, Bardet and Kammoun (2008) assures that the DFA method is not robust and should not be applied for trended processes or in case of changes of the mean.

Besides, several measures of entropy have been used to estimate the complexity of time series such as the Lyapunov exponents, Kolmogorov entropy and the permutation entropy. But one of the most used statistic in clinical application is the Approximate entropy (ApEn). It quantifies the unpredictability of a time series by computing the probability that patterns (or subsequence) of observations will repeat themselves within predetermined tolerance limits on the next incremental comparisons. ApEn reflects the likelihood that “similar” patterns of observations will not be followed by additional “similar” observations.

2.2 Mathematical modeling

Our point of view differs from the preceding approaches as being more theoretical: We prefer to base our statistical analysis on probabilistic modelling, see introduction.

The mathematical translation of a random signal with constant structural parameters is furnished by the notion of stationary process. This kind of process is well known in probability and widely used in engineering since the 1970's. Their main feature is the existence of a spectral representation (see Cràmer and Leadbetter, 1967). However, in this study, we restrict ourselves to Gaussian process and recall that every zero mean stationary Gaussian process admits a harmonizable representation (Yaglom, 1958)

$$X(t) = \int_{\mathbb{R}} e^{it\xi} f^{1/2}(\xi) W(d\xi), \quad \text{for all } t \in \mathbb{R}, \quad (2.1)$$

where $W(d\xi)$ is a complex Brownian measure chosen such that X is real valued, and f is an even function called the power spectral density. By adding the mean value μ , we shift to the case of stationary Gaussian process and expression 2.1 becomes:

$$X(t) = \mu + \int_{\mathbb{R}} e^{it\xi} f^{1/2}(\xi) W(d\xi), \quad \text{for all } t \in \mathbb{R}, \quad (2.2)$$

Roughly speaking, the power spectral density corresponds to Fourier Transform of the correlation structure. In this framework, the energy in a given frequency band $[\omega_1, \omega_2]$ corresponds to the quantity $\int_{\omega_1}^{\omega_2} f(\xi) d\xi$. More precisely, $LF = \int_{0.04}^{0.15} f(\xi) d\xi$ and $HF = \int_{0.15}^{0.5} f(\xi) d\xi$. Moreover, these quantities can be estimated as the modulus of Fourier transform of the process X multiplied by the inverse Fourier transform of the rectangular spectral window localized on frequency band $(0.04 Hz, 0.15 Hz)$, resp. $(0.15 Hz, 0.4 Hz)$. However, using a rectangular spectral window induces a bad time localization. For this reason, in the following, we prefer to use wavelet transform rather than Fourier transform.

Fractal index

In this subsection, we precise the relationship between so-called fractal index and the ratio LF/HF. It is well documented that the spectral density associated to heartbeat series satisfies power law on the HF, LF and VLF frequency bands, that is $f(\xi) = \sigma^2 \times |\xi|^{-\beta}$ where β is a positive real number. By using a double logarithm scale, this is equivalent to say that the log spectral density is an affine function of the log frequency:

$$\ln f(\xi) = -\beta \ln |\xi| + \ln \sigma^2 \quad (2.3)$$

for $\xi \in (0.15 Hz, 0.4 Hz)$ or $\xi \in (0.04 Hz, 0.15 Hz)$.

Next, for a Gaussian stationary process with representation (2.2), we can interpret as a fractal index the slope of the spectral density (in double log scale), see Bardet & Bertrand (2010) or Ayache & Bertrand (2011). In this double log scale the quantity $\ln(LF/HF) = \ln(LF) - \ln(HF)$ correspond to the slope of the spectral density multiplied by the distance between the average of LF frequency band and the average of HF frequency band. Thus, if the spectral density were an affine function (in double log scale), then fractal index and the ratio LF/HF were measuring the same quantity, up to a multiplicative constant. Unfortunately, spectral density of heartbeat series is not affine but is piecewise affine (in double log scale), see Huikuri et al.(2009), Cysarz et al.(2008) or Bardet & Bertrand (2010). Therefore, more theoretical studies are needed.

2.3 Dynamic approach for in vivo tachogram analysis

Due to the increase of life expectancy and the decrease of birth rate, developed and developing countries face the problem of ageing societies. The question is how to live longer and well without the need to

spent the end of one's life in a hospital or a center of hospitality in which not only living conditions are not so joyful and pleasant but also because of high cost health cares. Currently, the challenge is the development of monitoring systems for elderly people in ecological conditions. For instance, continuous monitoring may be of physical activities, physiological functions or daily habits in order to estimate a health condition. In case of danger, this will permit the intervention at the right time by the old person itself or an other person who may be more or less close in distance such as a relative, a doctor or a nurse via smart devices.

Thus, the aim of this work is to detect change of the structural parameters of a heartbeat time series that would mention a change in behaviour and so in biological functions.

More precisely, we want to detect changes on the regulation of the heartbeats, the regulation is measured through power spectral energy in the HF and LF bands. This ecological real time analysis may be interesting in other cases such as athletes or in case of some pathologies needing continuous monitoring like neuropathologies, for instance autism or epilepsy. So, a prediction of potential crises or a loss of control may prevent it and facilitate a better comprehension of the pathology by the concerned person or the others.

To this end, the theoretical framework used is that provided by the notion of locally stationary process (Dahlaus *et al.* 1997). To avoid technicality, a signal X is locally stationary if there exists time segmentation $\tau = \tau_1, \dots, \tau_{K+1}$, such that X is stationary on each interval (τ_k, τ_{k+1}) . K is the number of change points, and by convention the case $K = 0$ corresponds to stationary process with zero change points. For such locally Gaussian stationary process, we have the following spectral representation theorem

$$X(t) = \mu(t) + \int_{\mathbb{R}} e^{it\xi} f^{1/2}(t, \xi) dW(\xi), \quad \text{for all } t \in \mathbb{R}, \quad (2.4)$$

where

- the spectral density $f(t, \xi)$ is a piecewise constant function of time, i.e., there exists a partition τ_1, \dots, τ_K such that $f(t, \xi) = f_k(\xi)$ for $t \in [\tau_i, \tau_{i+1}[$
- the mean RR interval $\mu(t)$ is also piecewise constant for another partition $\tilde{\tau}_1, \dots, \tilde{\tau}_L$ with $\mu(t) = \mu_\ell$ if $t \in [\tilde{\tau}_\ell, \tilde{\tau}_{\ell+1})$.

In this framework, we are concerned by a problem of change point analysis. More precisely, we are concerned by two change points problems: detecting change on the mean and detecting change on power spectral density.

2.4 A new approach

For heartbeat time series, changes on the mean are easy to detect, when they are large enough. Moreover, change point analysis on the mean is a well documented problem, see the monographs Basseville & Nikiforov (1993), or Broadsky & Darkhovsky (1993). However, fast change point analysis with large or huge dataset was a challenging problem in the beginning of 21st century, let us refer to the introduction of Bertrand, Fhima and Guillin (2011) and the references therein.

On the other hand, change on the spectral density corresponds to change on hidden structural parameter, so they cannot be detected by eyes. Moreover there exist few references on the detection of changes on the spectral density, see Bardet and Kammoun (2007) or Ludena and Lavielle (1999). The first natural idea is to localize the static estimation procedure and then to test the existence of change

points. Localization of the spectral energy is obtained through replacement of Fourier transform by wavelet transform. Indeed, wavelet are well localized both in time and in frequency and they provide well-suited tool for processing signal with a time varying frequency behaviour see Mallat (1998), Bertrand *et al.* (2009), or Ayache & Bertrand (2011).

Recall on wavelet coefficient

Let ψ a mother wavelet with a good localization in time and frequency. Stress that there exists no function (except 0) with both finite time support and finite frequency support, but we can found mother wavelet with 0.999% of its $L^2(\mathbb{R})$ norm inside finite support, both in time and frequency. Roughly speaking, we can assume that the mother wavelet ψ has a time support $(0, L)$ and a frequency support (Λ_1, Λ_2) . Then by modulation/scaling, we can deduce two wavelets ψ_1 and ψ_2 with their frequency support include in the orthosympathetic, resp. parasympathetic frequency band, see Bertrand *et al.* (2009) or Ayache & Bertrand (2011).

Next, for every shift $b \in \mathbb{R}$, we define the associated continuous wavelet coefficient by

$$W_{\psi_j}(b) = \int_{\mathbb{R}} \psi_j(t-b)X(t) dt \quad \text{unit in } \textit{second}^2. \quad (2.5)$$

for $j = 1$ or 2 . Then, from Formula (2.4), we can deduce the following harmonizable representation of wavelet coefficients (see Abry et al 2010, Bardet & Bertrand 2010 or Ayache & Bertrand 2011) for all $(b, b+L_j) \subset (\tau_k, \tau_{k+1})$

$$W_j(b) = \int_{\mathbb{R}} e^{ib\xi} \widehat{\psi_j}(\xi) f_k^{1/2}(\xi) dW(\xi)$$

which turns to be a zero mean complex valued Gaussian process with variance $\mathbb{E}(|W_j(b)|^2) = \int_{\mathbb{R}} |\widehat{\psi_j}(\xi)|^2 f_k(\xi) d\xi$ where $|z|$ denotes the modulus of the complex number z . Thus we have

$$\log(|W_j(b)|^2) = \log\left(\int_{\mathbb{R}} |\widehat{\psi_j}(\xi)|^2 f_k(\xi) d\xi\right) + \zeta(b) \quad (2.6)$$

where $\zeta(b)$ is a random variable with law $\ln \chi^2$, for all b . Stress that in Formula (2.6), the modulus of the Fourier transform of the wavelet $|\widehat{\psi_j}(\xi)|^2$ appears as a spectral window.

Change-point problem

The transposition from static to dynamic study, leads us to consider the change point problem of the mean of the multivariate time series $Z_1(b)$, $Z_2(b)$ where

$$Z_1(b) = \log(|W_1(b)|^2), \quad \text{and} \quad Z_2(b) = \log(|W_2(b)|^2).$$

A huge literature exists on change point problem, see the monograph Basseville& Nikiforov (1993), Broadsky & Dharkovsky (1993), Montgomery (1997), Csorgo-Horvarth (1997) or the model selection approach in Birgé & Massart (2007). But in this study, we prefer to use a simple method well suited for large datasets, that is the Filtered Derivative with p-value (FDpV) method, Bertrand, Fhima and Guillin (2011). To put it into a nutshell, FDpV is a two step procedure, in which a so-called filtered derivative method is applied first and then a test is carried out to remove the false detection from the list of change points found in Step 1. This method detect change with a precision similar to other

method, its main advantage is its complexity: Both time and memory complexity of FDpV method are of order $\mathcal{O}(n)$, where n denotes the size of the dataset. Recall that Least Squares method with or without penalization are of order $\mathcal{O}(n^2)$, and that for a marathon runner $n \simeq 40,000$ and for a day in the life of a shift worker $n \simeq 120,000$.

Summary

To sum up, for modeling the heartbeat series, we choose a locally stationary gaussian process. We then use the continuous wavelet transform for calculating the spectral density energy in the HF and LF bands. Finally, for detecting changes on heart regulation, we adopt the FDpV segmentation.

3 Results

In this section, we give result of FDpV segmentation on heart rate and on log-wavelet energy into HF and LF bands for marathon runners and shift workers. Then, we present the difference between usual conditions and effort conditions, on the two cohorts, namely marathon runners and shift workers. In subsection 3.1 we present a paradigmatic example of a marathon runner, whereas in subsection 3.2 we present an example of a shift worker. Eventually, subsection 3.3 provide analysis of log energy effort versus resting for the two cohorts.

3.1 Results for marathon runners

We have the following raw dataset corresponding to B1 at Paris Marathon 2006. In Fig. 3.1, we can

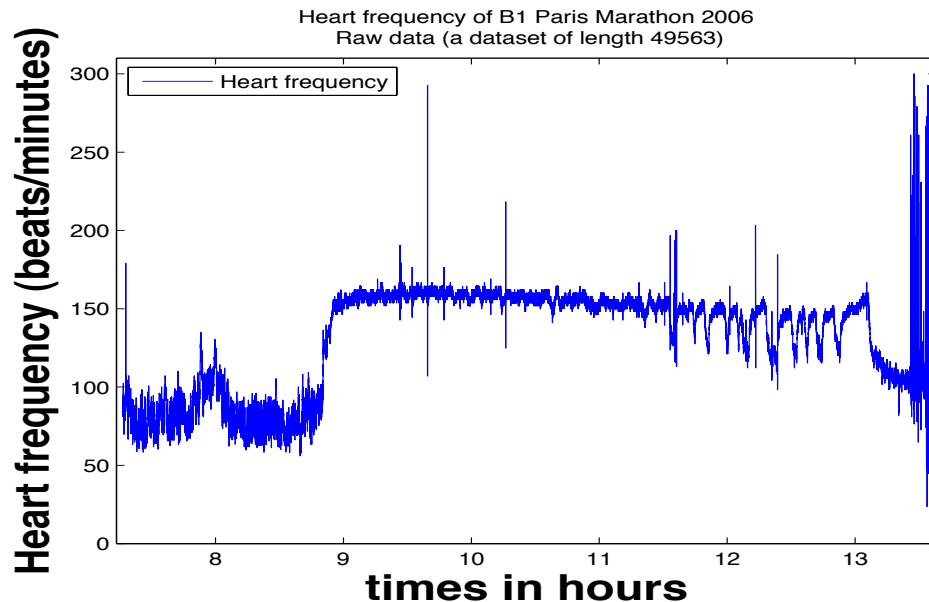
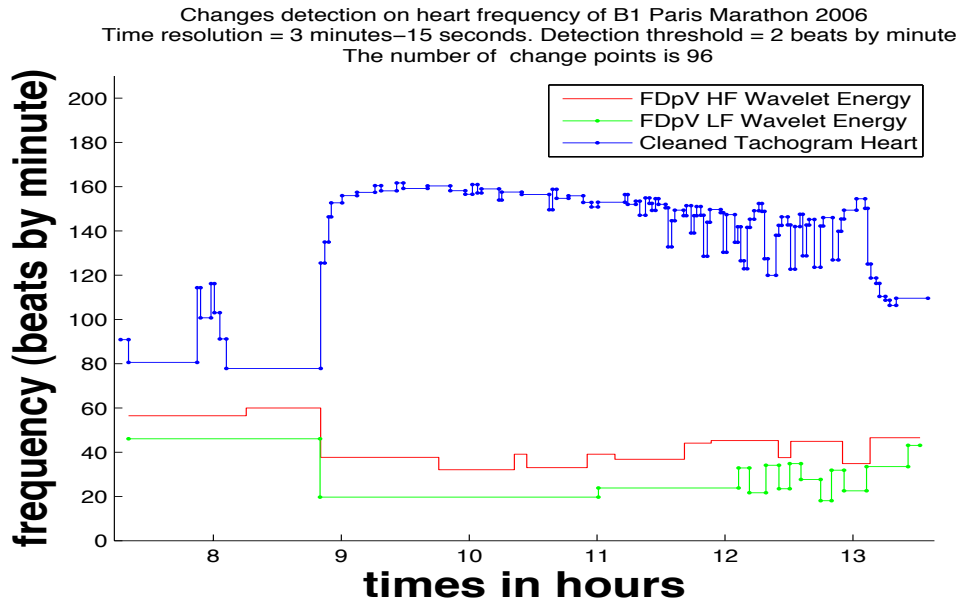


Figure 1: A raw HR time series of a marathon runner B1, Paris Marathon 2006 (49,000 heartbeats)

denote some aberrant data: For instance, human heart rate can not be larger than 220 beats/minute. These aberrant data are cleaned by using "TACHOGRAM CLEANING", a code developed by Nadia Khalfa

and Pierre R. Bertrand at INRIA Saclay in 2009. "TACHOGRAM CLEANING" corrects aberrant data by referring to physiological arguments rather than statistical ones. Then, heart rate is segmented by FDpV method. Actually, we segment the RR interval time series by FDpV method and then we translate the results into HR's corresponding ones, by using the equation $HR = 60/RR \text{ Interval}$. We also calculate the wavelet coefficient in HF and LF bands, then the corresponding log-wavelet energies, and finally, we use FDpV method to segment them. Fig. 2 is obtained by using "In vivo Tachogram Analysis (In ViTA)", a software developed by Pierre R. Bertrand at INRIA Saclay in 2010.



B1, compressed data, Paris Marathon 2006 (40,000 heartbeats)

Blue: Heart rate cleaned by "Tachogram cleaning" and compressed by FDpV.

Red: LF energy compressed by FDpV; Green: LF energy compressed by FDpV

Figure 2: HR regulation of the marathon runner in LF and HF bands

Let us comment Fig. 2: The blue line represents the segmented cleaned RR time series. The signal is compressed from around 40,000 heartbeat to 100 points. The compressed signal detects well the changes namely those occurring at the beginning of the race (just before 9 hours) by an abrupt increase and at the end of the race by an abrupt decrease (after 13 hours). The beginning of the race is equally very well detected in HF (green line) and LF (red line) energy by an abrupt decrease. Other changes are detected throughout the race. A notable increase in the HF energy is detected at the end of the race.

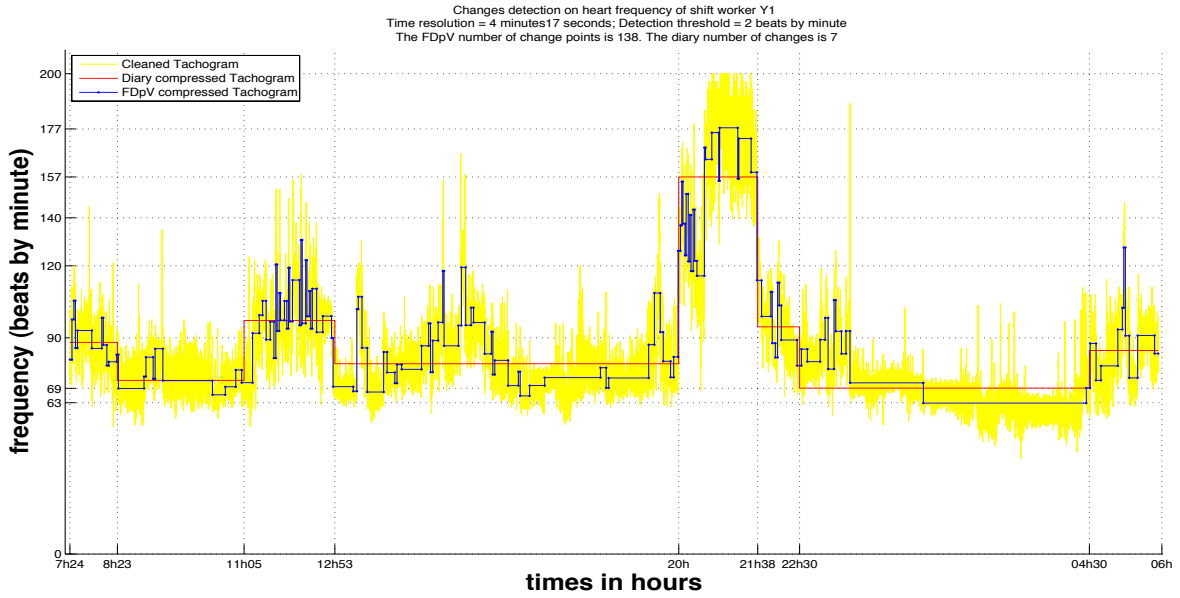
3.2 Results for shift workers of the company Z

We present an example of application to a daylong heartbeat time series of a shift worker in ecological situation. The shift worker Y1 has manually reported changes of activity on a diary, as shown in Table1. This dataset is first preprocessed using 'Tachogram cleaning' to avoid the maximum of artifacts and noise and then processed by "In vivo Tachogram Analysis (In ViTA)", to calculate wavelet coefficient in LF and HF bands and to apply FDpV segmentation. The computation time is around 1 minute 30 seconds

Time	7h24-8h23	8h23-11h05	11h05-12h53	12h53-20h	20h-21h38	22h30-4h30
Activity	Task 1	Task 2	Picking	Free afternoon	playing football	Sleeping

Table 1: A shift worker's activities record

for a dataset of size $n = 120,000$ with a code written in Matlab and a 2.8 GHz processor. This is faster than manual segmentation. Moreover, the FDpV segmentation is more accurate than the manual one. For instance, Y1 has reported "*Football training*" from 20h to 21h38, which is supported by our analysis. Indeed, with FDpV method, we can see more details of the training namely the warming-up, the time for coach's recommendations, and the football game with two small breaks.



A day in the life of Y1, shift worker (120,000 heartbeats)

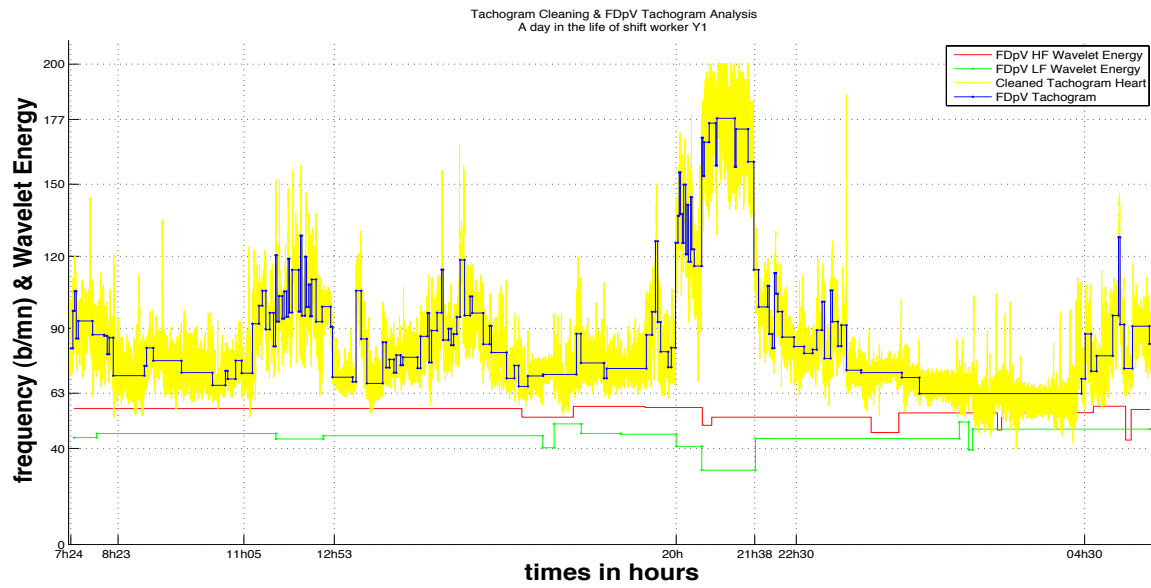
Yellow: Heart rate cleaned by "Tachogram cleaning".

Red: heart rate segmented by using the diary noticed by Y1.

Blue: Heart Rate automatically segmented and compressed by FDpV

Figure 3: A day in the life of a shift worker (120,000 heartbeats) segmented and compressed by FDpV

Note the decrease of HF wavelet energy during the football game.



A day in the life of Y1, shift worker (120,000 heartbeats)

Yellow: Heart rate cleaned by "Tachogram cleaning".

Blue: Heart Rate automatically segmented and compressed by FDpV

Red: Heart rate regulation in LF band

Green: Heart rate regulation in LF band

Figure 4: A day in the life of a shift worker (120,000 heartbeats) segmented and compressed by FDpV

3.3 Cohort study of wavelet energy

On Fig. 3 and Fig. 4, we can denote a variation of HF and LF energy following the activity. This led us to represent the activity by a point in the plane of HF and LF energy. More precisely, we use log-log coordinates where the x-axis corresponds to HF wavelet energy and the y-axis to LF wavelet energy.

The characteristic values related to marathon runners are gathered in Table 2 and presented in Fig.5 where each green point represents the maximum of wavelet energy reached by a runner before the beginning of the race, while each red point represents the minimum of wavelet energy reached during the race.

The distance between any point and the straight line \mathcal{D}_1 of equation $(y = x)$ or equivalently $\log LF = \log HF$ is the quantity $(y - x)$ or equivalently $\log(LF/HF)$.

As we can notice in Fig. 5, the (LF,HF) points corresponding to effort condition (marathon race in red) are clearly separated from the (LF,HF) points corresponding to so-called resting condition before the race (green points). The points are differentiated following an axis almost parallel to the first bissectrice \mathcal{D}_1 with equation $(y = x)$. Therefore the quantity $(y - x)$ or equivalently $\log(LF/HF)$ remains almost constant for these two sets of points. In conclusion, $\log(LF/HF)$ does not provide any relevant information, as previously noted for instance by Cysarz *et al.* (2008).

Marathon runners	Max Log HF	Max Log LF	Min Log HF	Min Log LF
1	12.4	14.4	5	9
2	12.4	15.6	5.4	9
3	14	15.6	4	8
4	9,5	12.56	4.32	8,32
5	11.3	13.7	5,5	9
6	11.5	15	4.5	8
7	10.8	14.2	5	7.7

Table 2: Log LF and HF of Marathon runners

Shift workers	Max Log HF	Max Log LF	Min Log HF	Min Log LF
1	12.7	14.7	7.7	10,9
2	11.9	14.5	8	12
3	14.5	16.2	8.9	12,8
4	12.9	16.2	9.8	11.3
5	13.2	15.2	9.4	10.9

Table 3: Log LF and HF of Shift workers

Next, we can use the same method for shift workers. Their characteristic values are gathered in Table 3 and plotted in Fig.6. Let us point a slight difficulty: For marathon runners, we denote a clear difference between physical activity and normal condition. For shift workers, sport activity was not included in the experimental protocol. Fortunately, we have found five shift workers with physical activity outside of working time. The better example is shift worker Y1, who reported football training from 8 p.m to 9 p.m 38 min. We can notice the same difference between effort conditions (sport) and normal condition. Eventually, we can observe that there exist no difference between marathon runners and shift workers

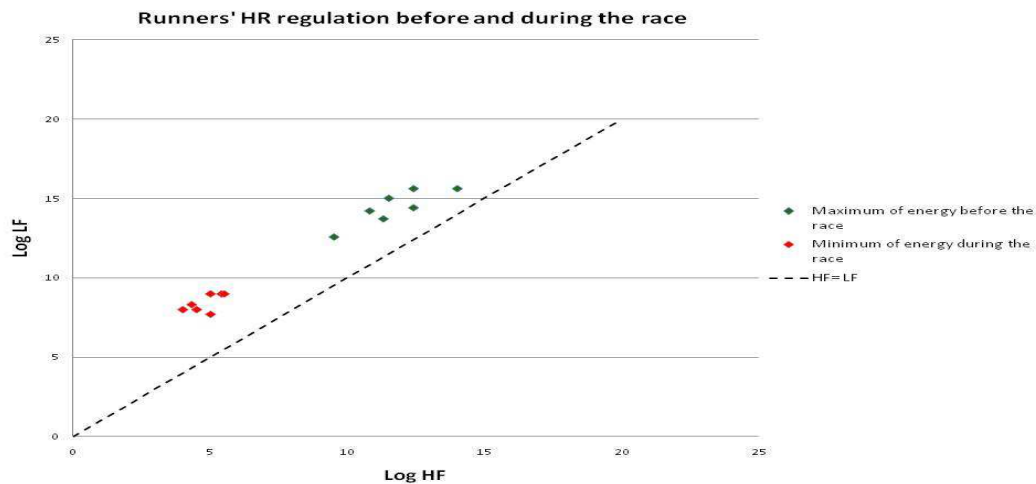


Figure 5: Runners' HR regulation in HF and LF bands



Figure 6: Shift workers' HR regulation in HF and LF bands

during normal or resting condition, but there exists a difference between sport condition as football game and marathon race.

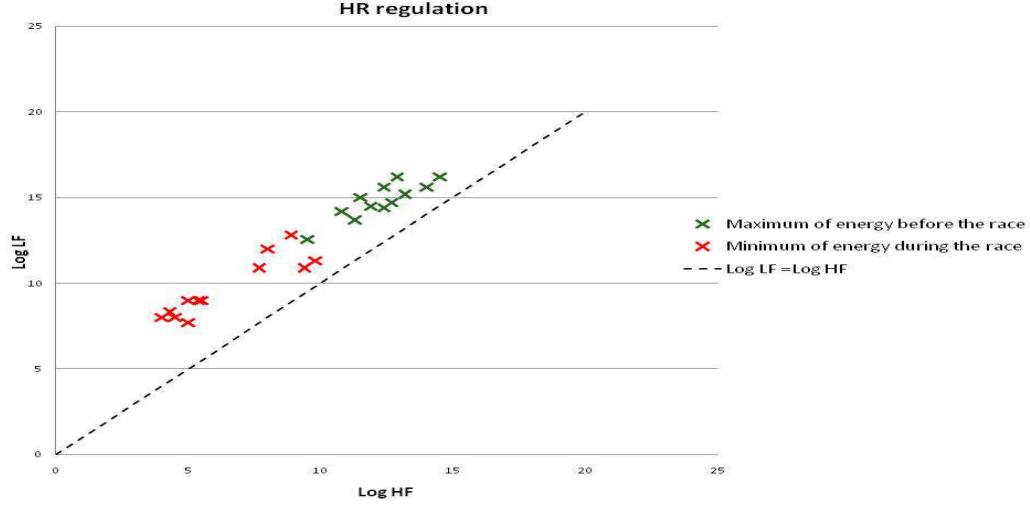


Figure 7: Runners' HR regulation in HF and LF bands

Statistical study of the cohorts

The change in the characteristic values of the ratio $\text{Log}(\text{LF}/\text{HF})$ before the beginning of the race and during the race i.e ($\text{Max Log LF} - \text{Max Log HF}$) and ($\text{Min Log LF} - \text{Min Log HF}$) were analyzed using a T-test of student for paired data ($\alpha = 0.05$) with R software. Similarly, variations in the characteristic values of the quantity ($\text{Log LF} + \text{Log HF}$) i.e ($\text{Max Log LF} + \text{Max Log HF}$) and ($\text{Min Log LF} + \text{Min Log HF}$) respectively before and during the race were analyzed using the same T-test.

The characteristic values of shift workers during rest and 'extreme' effort were analyzed following the same procedure.

Results

The difference between $\text{Log}(\text{LF}/\text{HF})$ before the beginning of the race and $\text{Log}(\text{LF}/\text{HF})$ during the race is not significant (p-value = 0.08, (Mean = 2.74, SD = 0.74) and (Mean = 3.61, SD = 0.47) respectively). Equivalently, for shift workers, this ratio is not different in rest and during high effort (p-value = 0.53, (Mean = 2.32, SD = 0.64) and (Mean = 2.82, SD = 1.23) respectively). While the quantity ($\text{Log LF} + \text{Log HF}$) is significantly different between the beginning of the race (Mean = 26.13, SD = 2.41) and the 'extreme moment' of the race (Mean= 13.25, SD= 1.02) and even for those of shift workers with (Mean= 28.4, SD= 1.64) and (Mean= 20.34, SD= 1.18) with p-values that are well under α and respectively equal to 1.310^{-5} and 6.110^{-5} .

These results come to support visual conclusions already obtained from Fig.5 and Fig.6. Indeed, the quantity ($\text{Log HF} + \text{Log LF}$) increases following the first bisectrice \mathcal{D}_1 with equation ($x = y$). Following this direction is parallel the points are well differentiated while the projection following the second bisectrice corresponding to the increase of quantity ($\text{Log LF} - \text{Log HF}$) or equivalently $\text{Log}(\text{LF}/\text{HF})$ is extremely narrow, see Fig.8.

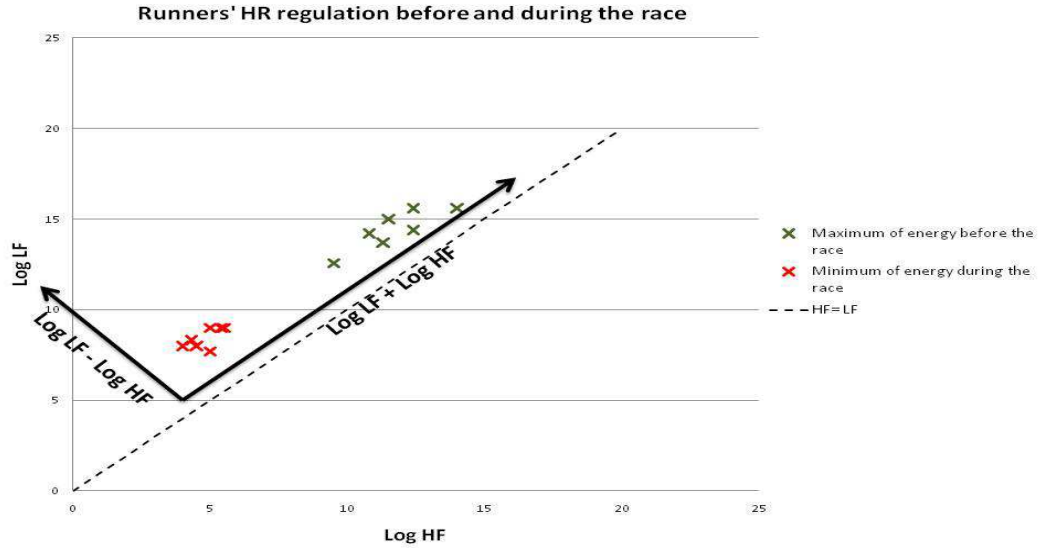


Figure 8: HR regulation following the axis x-y and x+y

4 Discussion

The juxtaposition of the data measured at rest of two categories of healthy subjects (sportsmen and working people) shows the same behaviour while activities of moderate intensity (professional context) and vigorous activities (sport context) can be ordered in a logical continuum.

The interpretation of data in this form has already opened new perspectives, complementing the classical scheme of the activation of the cardio-vascular system which is gradually resulted from the reduction of the parasympathetic influence then the stimulation of the orthosympathetic system. It appears here on the example of a marathon runner that the two components in the HF and LF bands are strongly correlated with concomitant, parallel and fast changes. The so-called neuro vegetative regulation might well have been described in other ways and according to a chronology much finer representing an instant rather than a state.

Other measures, in different circumstances (sleep, wakefulness, physical overtraining, burnout, drug treatment with agonists and antagonists of the autonomic nervous system) and on subjects with clearly defined characteristics (fitness, age, gender, cardiac, neurological or psychiatric disease) are needed to advance the understanding and interpretation of this new informations.

5 Conclusion

In conclusion, in the present study we calculate wavelet energies in LF and HF bands combined with the FDpV method for segmentation of the heartbeat series. On one hand, it compresses the heartbeat signal by a factor of 1000 with keeping important details and detecting very finite ones that can not be obvious visually (see the shift worker's example above). On the other hand, it permits to follow up the wavelet energy of the heartbeat signal in low and high frequency bands. In light of this, we were able to

reconsider the classical parameter LF/HF and opt for a new one in order to assess HR regulation during physical effort.

This work concerns one signal analysis; in further studies, we will investigate a multi signal analysis and focus on correlations for a better understanding and more precise detection of abrupt changes in biological and/or physiological mechanisms leading to change in behaviour that are manifested through a change in the attitude (epilepsy crises) or in the feelings and emotions (stress, nervousity). By taking into account the environment of work (in the case of shift workers for example) we can introduce psycho-physiological constraints that will provide additional and useful informations.

Acknowledgments: This work is part of the project "PHYSIOSTAT" (2009-2011) which is supported by device's grants from Digitéo and Région Île-de-France.

References

- [1] Ayache, A. and Bertrand, P.R. (2011), "Discretization error of wavelet coefficient for fractal like process", to appear in *Advances in Pure and Applied Mathematics*.
- [2] Bardet, J.M. and Bertrand, P.R. (2010), "A nonparametric estimator of the spectral density of a continuous-time Gaussian process observed at random times", *Scand. J. of Statistics*. Vol. 37, 458–476.
- [3] Bardet, J.-M. and Kammoun, I. (2008) Detecting abrupt changes of the long-range dependence for the self-similarity of a Gaussian processes. *C. R. Math. Acad. Sci. Paris* 346, no. 13-14, 789–794.
- [4] Bardet, J.-M. and Kammoun, I; (2008) Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes. *IEEE Trans. Inform. Theory* 54, N. 5, 2041–2052.
- [5] M. Basseville & I. Nikiforov, (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs, NJ.
- [6] Bertrand, P.R., Fhima, M. and Guillin, A. (2011) "Off-line detection of multiple change points with the Filtered Derivative with p-Value method." to appear in *Sequential Analysis*,
- [7] Bertrand, P.R., Teyssière, G., Boudet, G. and Chamoux, A. (2009). "Detection of Change-Points in the Spectral Density. With Applications to ECG Data", proceedings, EGC'09, Strasbourg, 28-30 janvier 2009. <http://hal.archives-ouvertes.fr/hal-00354416/fr/>.
- [8] Billat, V., Mille-Hamard L., Meyer Y. and Wesfreid, E. (2009). Scaling in heart rate and speed dynamics of runners in the marathon race. *Physica A*.
- [9] Boudet, G., Albuissou, E., Bedu, M. and Chamoux, A. (2004). Heart rate running speed relationships during exhaustive bouts in the laboratory. *Can. J. Appl. Physiol.* **29**(6), 731–742.
- [10] Chamoux, A. (1984). Le système Holter en pratique. *Médecine du Sport* **58**, 43-273, 54–284.
- [11] Chen X., Mukkamala R.(2007) Selective Quantification of the Cardiac Sympathetic and Parasympathetic Nervous Systems by Multi-Signal Analysis of Cardio-Respiratory Variability, *Am J Physiol Heart Circ Physiol*, 294, p. 362-371
- [12] Cottin, F., Lepêtre, P.M., Lopes, P., Papelier, Y., Médigue, C. and Billat V. (2006). Assessment of ventilatory thresholds from heart rate variability in well-trained subjects during cycling, *Int. J. Sports Med.* **27**, 959–967.
- [13] Ching E. S. C., Tsang Y-K. (2007). Multifractality and scale invariance in human heartbeat dynamics *Physical Review* **E76**

- [14] Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. Wiley and Sons.
- [15] Cysarz D., von Bonin, D., Brachmann, P., Buetler, S. Edelhäuser, F., Laederach-Hofmann, K. and Heusser P. (2008). Day-to-night time differences in the relationship between cardiorespiratory coordination and heart rate variability. *Physiol. Meas.* **29**:1281-1291.
- [16] D'Addio G., Pinna G.D., Maestri R., Corbi G., Ferrara N., Rengo F.(2004 IEEE Conferences), Quantitative Poincare Plots Analysis Contains Relevant Information Related to Heart Rate Variability Dynamics of Normal and Pathological Subjects *Computers in Cardiology* ,31, p.457-460.
- [17] Hogue CW Jr, Domitrovich PP, Stein PK, Despotis GD, Re L, Schuessler RB, Kleiger RE, Rottman JN. 1998, RR interval dynamics before atrial fibrillation in patients after coronary artery bypass graft surgery, *Circulation*, 98, p.429-434.
- [18] Douglas PS, O'Toole ML, Hiller WD, Hackney K, Reichek N. (1987) Cardiac fatigue after prolonged exercise. *Circulation* 76:1206–1213.
- [19] Huikuri HV , Valkama JO ,Airaksinen KE , Seppanen T ,Kessler KM ,Takkunen JT and Myerburg RJ Frequency domain measures of heart rate variability before the onset of nonsustained and sustained ventricular tachycardia in patients with coronary artery disease (1993) *Circulation* **4**:1220-1228
- [20] Huikuri HV, Perkiömäki JS, Maestri R, Pinna GD.,(2009) Clinical impact of evaluation of cardiovascular control by novel methods of heart rate dynamics. *Philos Transact A Math Phys Eng Sci.*,367, p.1223-38.
- [21] Ivanov C., Amaral L., Goldberger A.L., Havlin S., Rosenblum M.G., Struzik Z. R., Stanley H. E.(1999). Multifractality in human heartbeat dynamics. *Nature* **399**.
- [22] Kammoun I. (2007) Modélisation et détection de ruptures des signaux physiologiques issus de compétitions d'endurance.
- [23] Ludena, C. and Lavielle, M. (2000) The multiple change-points problem for the spectral distribution. *Bernoulli*, 6, 845–869.
- [24] Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press.
- [25] Peng C.K., Buldyrev S., Havlin S.V., Simons M., Stanley H.E., Goldberger A.L., Mosaic organization of DNA nucleotides, 1994 *Phys. Rev.*, E 49, p.1685-1689
- [26] Picho V., Gaspoz J-M, Molliex S., Antoniadis A., Busso T., Roche F.,Costes F., Quintin L., Lacour J-R., Barthélémy J-C. (1999) Wavelet transform to quantify heart rate variability and to assess its instantaneous changes, *Journal of Applied Physiology*, 86, p. 1081-1091
- [27] Pikkujämsä S.M., Mäkikallio T.H., Airaksinen K.E., Huikuri H.V. (2001). Determinants and interindividual variation of R-R interval dynamics in healthy middle-aged subjects. *Am. J. Physiol. Heart Circ. Physiol.* 280(3) 1400–1406.
- [28] Singh N., Mironov D., Armstrong P W., Ross A M., Langer A.(1996) Heart Rate Variability Assessment Early After Acute Myocardial Infarction, Pathophysiological and Prognostic Correlates *Circulation*,93, p.1388-1395.
- [29] Task force of the European Soc. Cardiology and the North American Society of Pacing and Electrophysiology (1996), Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Circulation* 93 (5), 1043-1065.
- [30] Tsuboi K., Deguchi A., and Hagiwara H.(2010) Relationship between Heart Rate Variability using Lorenz Plot and Sleep Level, *32nd Annual International Conference of the IEEE EMBS*
- [31] Yaglom, A.M. (1958). Correlation theory of processes with stationary random increments of order n . *Trans. A.M.S.*, 8, p.87-141.